



Universidad
Carlos III de Madrid



Documento publicado en junio de 2017 en:

www.709mediaroom.es

Utray, F., Márquez, S. & Ochoa, L. (2017) “Sonido para vídeo UHD y cine 4K”, 709MediaRoom.

Sonido para vídeo UHD y cine 4K

Francisco Utray (UC3M), Sergio Márquez (UC3M) y Luis Ochoa (709 MediaRoom)

Publicado en junio de 2017 en www.709mediaroom.es

Resumen: El sonido es una pieza fundamental para la mejorar de la experiencia de usuario ante los contenidos audiovisuales. En este artículo hacemos un repaso de los estándares de calidad de sonido aplicables al video en ultra alta definición y el cine 4K

Comenzamos describiendo los parámetros de calidad del sonido digital: frecuencia de muestreo (Sample Rate) y profundidad de bits (Bit Depth). A continuación, abordamos una cuestión fundamental: la sonoridad (loudness) que nos acerca a conceptos como el rango dinámico del sonido y la compresión dinámica.

Una vez expuestas estas bases teóricas repasaremos los distintos códecs para sonido envolvente que se están utilizando o desarrollando para el cine 4K y el vídeo de ultra alta definición. Lo hemos organizado en cinco apartados: 1) Los inicios del audio digital para cine 2) Sonido inmersivo, sonido 3D o sonido basado en objetos 3) Sonido para salas de exhibición de cine 4) Códecs Sonido para televisión en UHD y Blu-ray 5) Códecs de sonido para difusión por internet

Palabras clave: Sonido envolvente, Ultra alta definición, UHD, 4K, cine Digital

Citar como: Utray, F.; Márquez S. & Ochoa, L. (2017) "Sonido para vídeo UHD y cine 4K", 709MediaRoom.

Sonido para UHD y 4K	3
Frecuencia de muestreo (Sample Rate) y profundidad de bits (Bit Depth)	4
Compresión dinámica / compresión de datos	8
Rango dinámico y sonoridad	10
Códecs para sonido envolvente	13
Códecs para sonido multicanal	13
Los inicios del audio digital para cine	15
Sonido inmersivo, sonido 3D o sonido basado en objetos	18
Sonido para salas de exhibición de cine	24
Códecs Sonido para televisión en UHD y Blu-ray	26
Códecs de sonido para difusión por internet	29
Conclusiones	31

Sonido para UHD y 4K

El sonido es una pieza fundamental para la mejorar de la experiencia de usuario ante los contenidos audiovisuales. Hemos entrevistado a [Sergio Márquez](#), de [Nómada Media](#), que generosamente ha compartido con nosotros sus conocimientos y nos ha ayudado a plasmar en esta sección las claves para el registro, postproducción y difusión del sonido para cine y el vídeo de ultra alta definición.



Vamos a comenzar describiendo los parámetros de calidad del sonido digital: [frecuencia de muestreo](#) (Sample Rate) y [profundidad de bits](#) (Bit Depth). A continuación, abordamos una cuestión fundamental: la [sonoridad](#) (loudness) que nos acerca a conceptos como el [rango dinámico](#) del sonido y la [compresión dinámica](#).

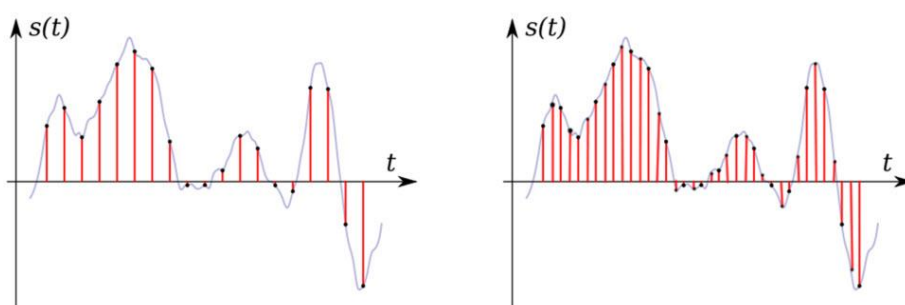
Una vez expuestas estas bases teóricas repasaremos los distintos [códecs](#) para [sonido envolvente](#) que se están utilizando o desarrollando para el cine 4K y el vídeo de ultra alta definición. Lo hemos organizado en cinco apartados:

- Los inicios del audio digital para cine
- Sonido inmersivo, sonido 3D o sonido basado en objetos
- Sonido para salas de exhibición de cine
- Códecs Sonido para televisión en UHD y Blu-ray
- Códecs de sonido para difusión por internet

Frecuencia de muestreo (Sample Rate) y profundidad de bits (Bit Depth)

Los dos parámetros técnicos que definen la calidad de una señal digital de sonido son la frecuencia de muestreo y la profundidad de bits.

La frecuencia de muestreo (Sample Rate) es el número de muestras que se toman por segundo para generar la señal digital. Se expresa con un valor de ciclos por segundo en kilohercios (kHz).



Representación de dos valores distintos de frecuencia de muestreo. El eje horizontal representa el tiempo y el vertical el valor de la señal de audio. Fuente: Progenerator.

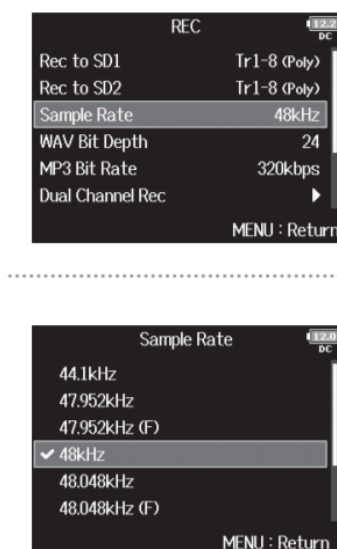
La frecuencia de muestreo que se utiliza para registrar y procesar el sonido viene determinada por la capacidad sensorial de los seres humanos: se considera que podemos escuchar los sonidos que se sitúan entre 20 y 20.000 Hz. Según el teorema de Nyquist, es necesario que la frecuencia de muestreo sea superior al doble de la máxima frecuencia a muestrear. Por lo tanto, la frecuencia de muestreo ha de ser como mínimo de 40 KHz.

Los valores de frecuencia de muestreo que se utilizan habitualmente para sonido profesional son las siguientes:

Frecuencia de muestreo	
Audio CD	44,1 kHz
Broadcast / DVD	48 kHz
DVD Audio / Blu-ray	96 kHz
Blu-ray	192 kHz

En producción musical para CD se ha utilizado tradicionalmente 44,1 kHz por la limitación de la capacidad de almacenamiento del soporte. Pero en producción audiovisual se considera insuficiente este valor de frecuencia de muestreo.

El estándar en producción audiovisual es 48 kHz. Las cámaras y los grabadores, desde la gama usuario hasta el nivel más profesional, utilizan esta frecuencia de muestreo. Es también el estándar para la difusión televisiva (broadcast) y para los discos DVD o Blu-ray. Igualmente, en captación de sonido directo, lo habitual es utilizar 48 kHz.



Menú de selección de la frecuencia de muestro de un grabador portátil de audio: Fuente: [Zoom](#)

En los estudios grabación y postproducción para audiovisual y cine de alta calidad, la tendencia es trabajar a 96 kHz. Aunque el material de sonido directo esté grabado a 48 kHz, el técnico hace la postproducción y la mezcla (reverbs, delays...) a 96 kHz. Este sobremuestreo aporta una serie de ventajas (mayor capacidad de alteración de un sonido original etc.) y permite dejar hasta el último momento la máxima calidad posible.

También en captación es cada vez más frecuente la grabación a 96 KHz, en tanto que los armónicos (múltiplos enteros que acompañan a cualquier frecuencia) son muy importantes para percibir con mayor naturalidad un sonido. Sobre todo, es fundamental el primer armónico (x2). Por ejemplo, cuando un violín emite una nota base de 15 kHz, su primer armónico es 30 kHz, y por lo tanto necesitamos 60 KHz de frecuencia de muestreo para poder capturarlo y reproducirlo. Por eso tiene sentido trabajar en un sistema completo de 96 KHz.

Las frecuencias de muestreo superiores, como 192 kHz, se utilizan en niveles 'audiófilos', en colecciones o sellos elitistas, como por ejemplo algunas grabaciones de música clásica de Deutsche Grammophon o en Jazz, las producciones de Chesky Records. Estas producciones intentan alcanzar la máxima calidad de sonido posible. Pero a nivel perceptivo, la mayoría de las personas no llegan a escuchar esta diferencia.

En teoría por lo menos, una frecuencia de muestro de 48 kHz, es suficiente para registrar eficazmente la capacidad auditiva del ser humano.

La profundidad de bits (Bit Depth o cuantificación) es el número de bits que se utilizan para registrar cada una de las muestras. Cuantos más bits se utilicen, más información queda registrada y el archivo resultante ocupará más espacio en el soporte de la grabación. Por lo tanto, tiene que ver con la precisión de la muestra de audio que se cuantifica y con cuanta información se define.

Los valores que se utilizan habitualmente en audio profesional son los siguientes:

Profundidad de bits (cuantificación)	
Audio CD	16 bits
Broadcast / DVD / Blu-ray	16 bits
Sonido de alta calidad	24 bits
Máxima calidad	32 bits
Postproducción máxima calidad	32 bit float

En captación de sonido directo casi todo se hace en 24 bits, aunque queda una herencia, todavía fuerte, de 16 bits. Los equipos ya están preparados desde hace años para 24 bits y pueden grabar sin problemas a 32 bits.

La profundidad de bits está íntimamente relacionada con el [rango dinámico de audio](#). Es decir, con la capacidad de registrar con calidad los sonidos más leves y los de nivel más alto. La percepción humana del sonido se estima en 120 decibelios de rango dinámico y con 24 bits es más que suficiente para registrar estos valores.

En postproducción se puede hacer todo a 32 bits, a 32 bits coma flotante (float) e incluso algunas tarjetas gráficas a valores superiores. Pero lo más habitual es trabajar a 24 bits.

La representación en coma flotante (en inglés floating point) indica que para los cálculos que realiza el equipo, se van a utilizar tantos decimales como sea necesario para garantizar la preservación de la calidad original. En los equipos de postproducción de alta calidad se puede elegir esta opción.

Estas dos variables, frecuencia de muestreo y profundidad de bits, dan como resultado un flujo de transferencia (BitRate) que se mide en bits por segundo (b/s, Kb/s o Mb/s).

Los valores de flujo de transferencia con los se suele trabajar en las distintas fases de la producción y difusión de sonido son los siguientes:

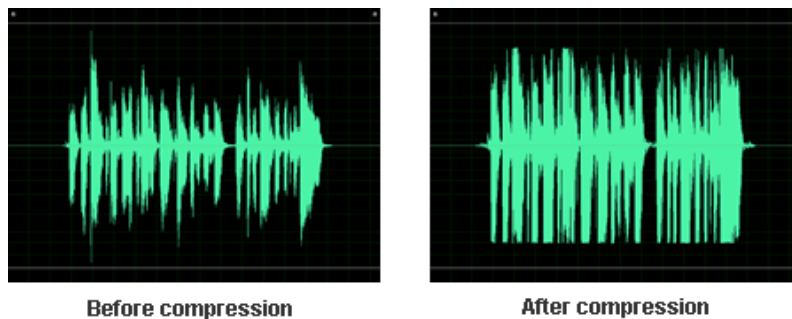
Flujo de transferencia	
Sonido mono comprimido MP3-64 (calidad insuficiente)	64 Kb/s
Audio estéreo con compresión AAC (calidad suficiente)	256 Kb/s
Dolby Digital 5.1 en DVD	448 Kb/s
Audio estéreo PCM sin compresión (16bits; 48khz)	1,5 Mb/s
Audio 5.1 PCM sin compresión (24bits; 48khz)	7 Mb/s
DTS HD Master Audio 7.1 (24bits; 96 khz)	24 Mb/s

Más adelante profundizamos en el concepto de [rango dinámico del sonido](#). Pero primero tenemos que repasar la diferencia entre [compresión dinámica](#) y [compresión de datos](#).

Compresión dinámica / compresión de datos

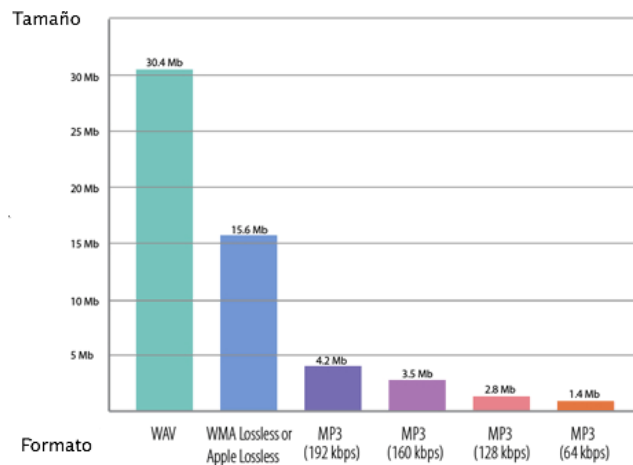
En sonido digital la palabra “compresión” se utiliza en dos contextos diferentes con significados muy distintos:

- 1) **Compresión dinámica.** En postproducción de sonido, la compresión es una manipulación de los niveles dinámicos de la señal que tiene como objetivo reforzar los sonidos más débiles y atenuar los más altos. Se utilizan los compresores dinámicos típicamente en radiodifusión y en música popular para aumentar la **sonoridad**. Las producciones de sonido de alta calidad evitan el uso de compresores en favor del **rango dinámico**.

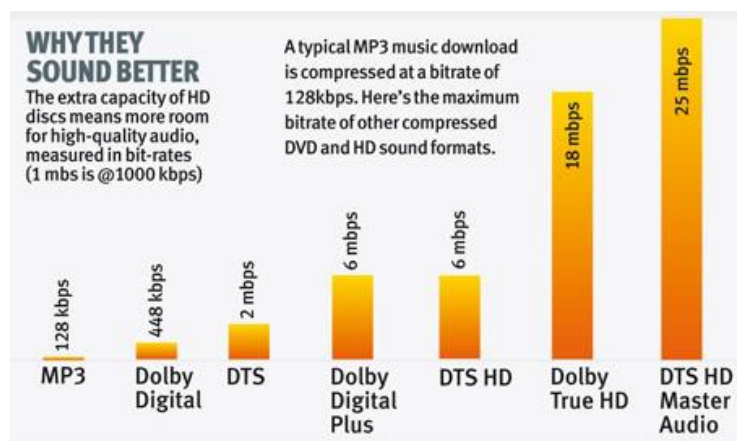


Compresión dinámica de una señal de audio. fuente: [Articulate](#)

- 2) **Compresión de datos.** Cuando se hace la codificación digital del sonido, la compresión es la reducción del tamaño de los archivos mediante unos algoritmos que llamamos ‘códecs de audio’. El ejemplo más conocido de compresión de datos para sonido es el formato ‘.mp3’, que reduce drásticamente el espacio que ocupa un archivo de audio en un disco duro, a costa de una pérdida de información y por lo tanto de la calidad. En difusión de sonido para cine la norma de compresión más extendida es ‘Dolby Digital’.



Compresión de datos MP3. fuente: Sarte Audio Elite



Compresión de datos en los formatos de máxima calidad. Fuente: The Media Server

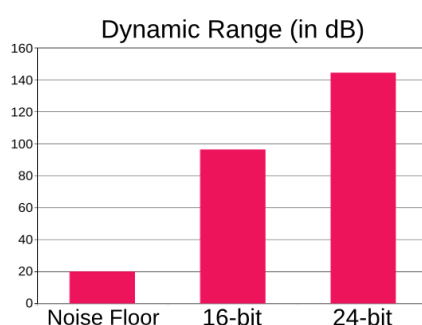
En la siguiente sección vamos a hablar de **sonoridad** y **rango dinámico** que son conceptos asociados a la compresión dinámica.

En el apartado dedicado a los **códecs de sonido envolvente** nos vamos a referir a la 'compresión de datos' en los estándares de sonido profesional de alta calidad.

Rango dinámico y sonoridad

El rango dinámico del sonido es la diferencia entre los valores de nivel más bajos y los más altos. Se mide en decibelios (db).

Existe una relación directa entre **profundidad de bits** y rango dinámico del sonido: cada bit que se emplea para generar la señal digital incrementa en 6 db la capacidad de registro de presión sonora. Con 16 bits se alcanzan 96 db de rango dinámico y se amplía a 144 db para 24 bits. Este valor de rango dinámico es, en teoría, suficiente, puesto que la capacidad del ser humano para percibir el sonido se sitúa en torno a los 120 db. Por encima de 120 db se encuentra el umbral del dolor.

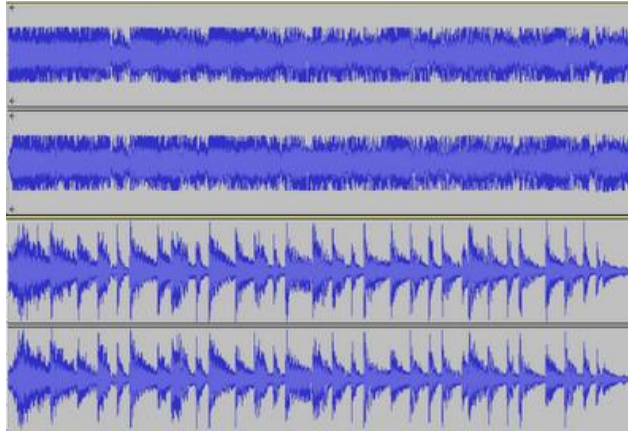


Capacidad de registro del rango dinámico en función de la profundidad de bit. Fuente: Libremusicproduction.

En grabaciones de música clásica tiene mucho sentido utilizar 24 bits, ya que la diferencia de dinámica entre un Piano-Pianissimo y toda la orquesta tocando a la vez es muy grande.

En cine, también es muy importante el rango dinámico. En estudios de grabación de orquestas para bandas sonoras suelen preferir utilizar 32 bits para capturar más dinámica. Las excelentes condiciones de reproducción del sonido en una sala de exhibición permiten apreciar con claridad tanto los sonidos más leves como el estruendo de las escenas de acción.

Sin embargo, en radio o en televisión, se utilizan compresores para reducir el rango dinámico y el resultado es que no hay tanta diferencia entre los niveles de sonido más bajos y los más altos. Esta **compresión dinámica** se hace para aumentar la sonoridad (loudness) y ayudar a la inteligibilidad del sonido en condiciones críticas de audición, por ejemplo, al escuchar la radio en un coche o la televisión en un ambiente ruidoso. El resultado es que tanto los susurros como los gritos quedan a un nivel alto e inteligible. La víctima en este proceso es el rango dinámico y el resultado se aleja de la percepción natural del sonido.



Forma de onda de la señal sin comprimir (abajo) y con compresión dinámica (arriba). Fuente: [Wachusett](#).

Esta cuestión se hace patente en la sensación de volumen que tiene la publicidad en televisión. Los radiodifusores se han dado cuenta de que las mediciones de los picos de sonido no reflejan la sensación de volumen de un programa. Los productores de los anuncios hacen las mezclas de sonido con mucha compresión dinámica, sacrificando el rango dinámico en favor de una mayor sonoridad. Esto genera quejas de los usuarios y baja los estándares de calidad de sonido:

“Todo suena super-plano, y se hace una guerra de decibelios para ver quien suena más fuerte” ([Sergio Márquez](#))

La normativa europea EBU R 128 (2011) “Normalización de la sonoridad y nivel máximo permitido de las señales de audio” trata de poner orden en esta cuestión:

“The switch from audio peak-normalization to loudness normalization is probably the biggest revolution in professional audio of the last decades. It is important for broadcasters to be aware of the loudness paradigm and how to adopt their systems and working practices accordingly.” ([Roger Miles, EBU](#))



“Lo importante no es quien suena más fuerte, sino que la película que están poniendo en televisión suene parecida a cómo se escuchaba en la sala de cine, es decir, con sus picos altos y sus pasajes de susurros. Tenemos la tecnología suficiente para tratar 120 db de rango dinámico. ¡Pues olvidémonos de estar sonando todo el rato a 100 db en TV! Procuremos respetar una dinámica apropiada para cada entorno.” ([Sergio Márquez](#))

“Si un compositor quiere transmitir ruido, agresividad, malestar, ira, etc.... parece buena idea hacer salvajadas con el sonido, pero si la razón es la de ‘es que no quiero sonar más bajo que los demás’, creo que es mejor que se vaya haciendo a la idea de que las reglas del juego han cambiado.” ([Ibon Larruzea](#))

Códecs para sonido envolvente

Códecs para sonido multicanal

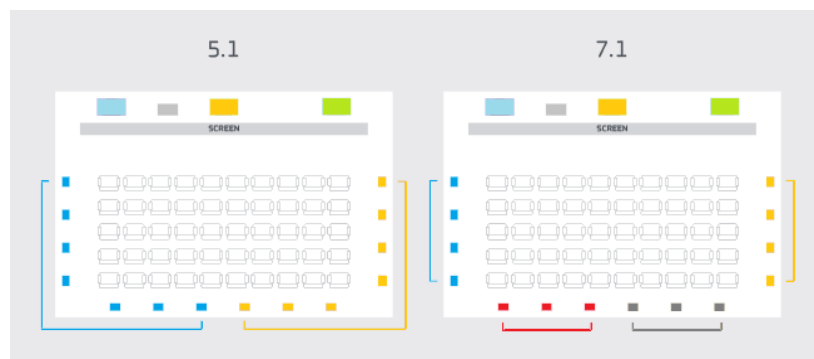
Fantasia (1940, Walt Disney) es la primera película que experimentó con el sonido envolvente (surround sound) para cine. Desde entonces se han desarrollado muchos sistemas de sonido para mejorar la experiencia de usuario tanto en las salas de exhibición como en el hogar (home cinema).



Colocación de los altavoces en un sistema multicanal de sonido envolvente para una sala de exhibición de cine. Fuente: wikipedia

En la actualidad, para cinematografía nos podemos encontrar con distintas configuraciones para sonido envolvente:

- L.C.R. (*left, center, right*): añade un canal central al clásico formato estéreo de la industria musical. Una variante, ya en desuso, es LCRS que utiliza un único canal surround.
- Surround 5.1: añade dos canales laterales (*surround left, surround right*) y un canal de subgraves (*subwoofer*).
- Surround 7.1: añade al 5.1 dos canales traseros (*rear left, rear right*).
- Sonido inmersivo, sonido 3D o sonido por objetos: es una nueva concepción del sonido envolvente que no está basada en el número de canales sino en el posicionamiento espacial de cada fuente de sonidos.



El sistema 7.1 dedica dos canales independientes para el sonido trasero izquierdo y derecho diferenciados de los canales laterales. Fuente: [Tech Review](#)

En esta sección vamos a hacer un repaso de los códecs para sonido envolvente que se utilizan en cine y difusión de vídeo de ultra alta definición. Está estructurada en cinco apartados:

- Los inicios del audio digital para cine
- Sonido inmersivo, sonido 3D o sonido basado en objetos
- Sonido para salas de exhibición de cine
- Códecs de sonido para televisión en UHD y Blu-ray
- Códecs de sonido para difusión por internet

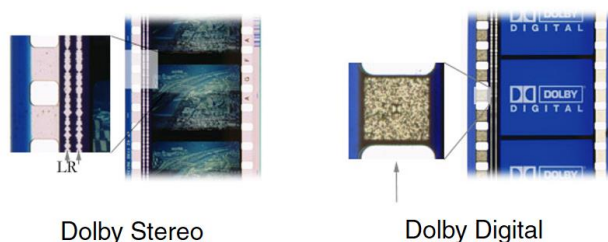
Los inicios del audio digital para cine

En los años previos a la proyección de cine digital (DCI) había tres sistemas de audio digital para salas de exhibición: Dolby, DTS y SDDS. Estos tres formatos solventaron las limitaciones técnicas para registrar, almacenar y procesar sonido digital; así como la sincronización del sonido digital con la proyección de película de 35 mm.

El primero en llegar en esta carrera fue 'Dolby Digital (AC-3)' que se estrenó en 1992 con *Batman Returns* (Tim Burton). Fue el primer sistema de audio multicanal digital 5.1 para cine. Era un códec con mucha pérdida de información, 10 a 1 (10:1), pero que discriminaba muy bien los sonidos más relevantes para el ser humano basado en su excelente codificación psicoacústica. Técnicamente era muy eficiente ya que lograba meter seis canales en 448 kb/s. Sin compresión, el equivalente serían casi 7 Mb/s.



La sincronización con la proyección se hacía imprimiendo las pistas de sonido en el celuloide. Esta técnica se llevaba usando desde los inicios del cine sonoro. En la siguiente imagen se puede ver la impresión de sonido analógico de 'Dolby Stereo', que se empezó a utilizar con *Star War* (1977, George Lucas), y la solución de Dolby Digital de los años 90 que empleaba el espacio de película entre las perforaciones para imprimir la señal de audio digital.



Impresión de las pistas de sonido en el celuloide: Dolby Stereo y Dolby Digital. fuente: Brian Florian

Justo después del lanzamiento de 'Dolby Digital', con *Jurassic Park* (1993, Steven Spielberg), apareció 'DTS Digital Theater System' que daba más calidad y utilizaba

mucha menos compresión (3:1). Este incremento de calidad gustaba sobre todo a los directores y al equipo técnico-artístico responsable de la banda sonora, pero era más complejo y costoso. DTS utilizaba un sistema propio para sincronizar el audio digital con la proyección de la película de 35 mm mediante un código de tiempo impreso en el celuloide y unos discos CD.



El tercer sistema, 'SDDS Sony Dynamic Digital Sound', se lanzó con Last Action Hero (1993, John McTiernan) y estaba instalado sobre todo en las salas de exhibición que tenían acuerdos de distribución con Sony. Esta patente definió un primer sistema 7.1, que es distinto al que actualmente se utiliza: era una configuración de canales 7.1, ubicando detrás de la pantalla cinco de ellos (Left, Left-Center, Center, Right-Center y Right). Solo las grandes producciones de Hollywood mezclaban en este formato (principalmente Columbia Pictures, filial de Sony) que, además, al necesitar más altavoces y amplificadores era mucho más caro de instalar para el exhibidor.



De estas tres soluciones para el audio digital en salas de cine, Dolby, que llegó primero, consiguió convertirse prácticamente en el estándar de facto.

Años más tarde, cuando se sustituyeron los proyectores de película 35 mm por los sistemas digitales DCI, la sincronización del sonido dejó de ser un problema. El soporte de las películas pasó a ser un disco duro DCP con capacidad suficiente para almacenar y reproducir sonido sin compresión y se pudieron usar directamente los archivos en formato PCM (.wav) tanto para estéreo como para sonido envolvente 5.1 o 7.1.

A partir de entonces, para la proyección de cine digital, se puede hacer sonido multicanal sin pagar ningún tipo de licencia, ni royalties de patente alguna.

En la siguiente captura de pantalla se puede ver el menú del software 'OpenDCP' para la creación de un archivo de audio 7.1 para salas de cine: se parte de 8 canales mono en formato PCM (.wav) sin comprimir para generar un archivo MXF.

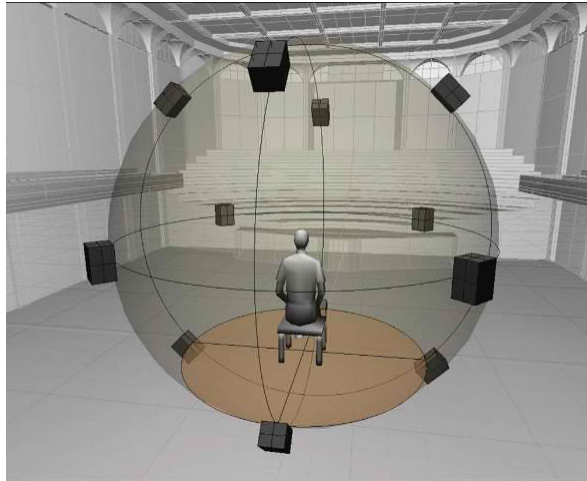
The screenshot shows the OpenDCP application window. At the top, there are tabs for 'JPEG2000', 'MXF' (which is selected), 'Subtitles', and 'DCP'. Below the tabs, the 'MXF Parameters' section includes a 'Type' dropdown set to 'WAV', an 'Enable Encryption' checkbox, a 'Label' dropdown set to 'SMPTE', a 'Key' text field with a long string of zeros, and a 'Frame Rate' dropdown set to '24'. Below this, the 'Sound Input Type' section has radio buttons for 'Mono' (selected) and 'Multi-Channel'. The 'Sound Output Parameters' section has radio buttons for 'Stereo', '5.1', and '7.1' (selected), along with a 'Hearing/Visually Impaired' checkbox. The 'Input Files' section contains ten rows, each with a label (Left, Right, Center, Sub, Left Surround, Right Surround, Left Center, Right Center, HI, VI-N) and a text field with a browse button. The 'Output File' section has a 'Sound' label and a text field with a browse button. A 'Create MXF' button is located at the bottom right of the interface.

Configuración de los canales de sonido para 7.1 para DCP con OpenDCP. El sistema crea un archivo MXF a partir de los ocho canales mono en formato PCM (.wav)

El siguiente apartado lo dedicamos al 'Sonido inmersivo, sonido 3D o sonido basado en objetos', que ha irrumpido con fuerza en el mundo del cine. El ejemplo más conocido es 'Dolby Atmos', pero hay también otros sistemas compitiendo en este mercado emergente.

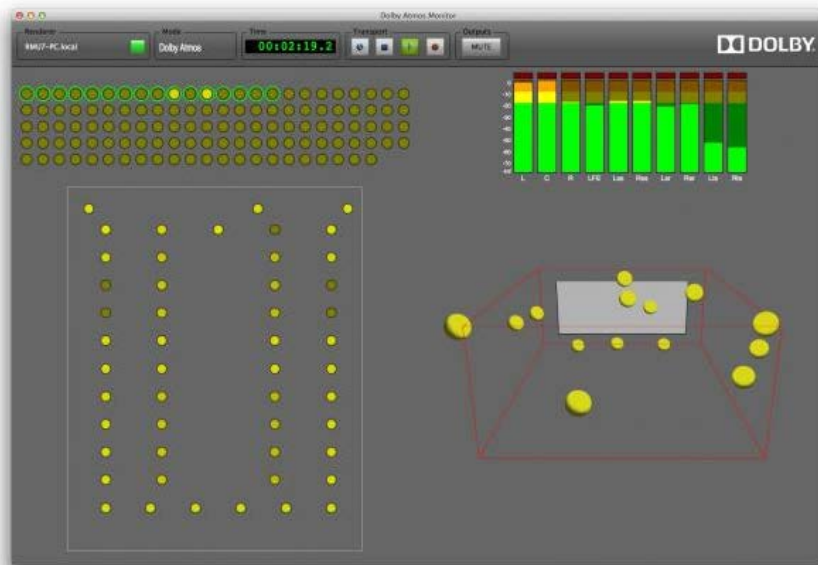
Sonido inmersivo, sonido 3D o sonido basado en objetos

El nuevo paradigma en audio es el sonido inmersivo, sonido 3D o sonido basado en objetos, que tiene aplicaciones en exhibición de cine en salas, en realidad virtual y también en el entorno de consumo doméstico de contenidos audiovisuales. El nuevo modelo ya no está basado en el número de canales: las mezclas para sonido inmersivo ubican cada fuente sonora, denominadas objetos, en el espacio 3D.



Esquema de la recepción de sonido inmersivo. Fuente: Newsbytes

El campo de sonido que percibe el ser humano es como una esfera y nosotros estamos dentro de ella escuchando distintas fuentes de sonido u objetos. Con los sistemas de sonido basados en canales, los creadores tienen que pensar en la configuración que tiene el cliente final (2.0, 5.1 o 7.1) para que los efectos de espacialidad suenen eficazmente. Pero los sistemas de audio basados en objetos ofrecen mayor libertad creativa porque se incluye en los metadatos una descripción espacial de la ubicación de cada objeto. A partir de estos metadatos, el procesador del audio tendrá que ajustar la mezcla en función del número de altavoces disponible y su posición en cada sala. El procesador determinará automáticamente cómo usar estos metadatos para crear el campo de sonido más inmersivo posible.



Interfaz de usuario de Dolby para la colocación de objetos sonoros en el espacio 3D. Fuente: [Dolby](#)



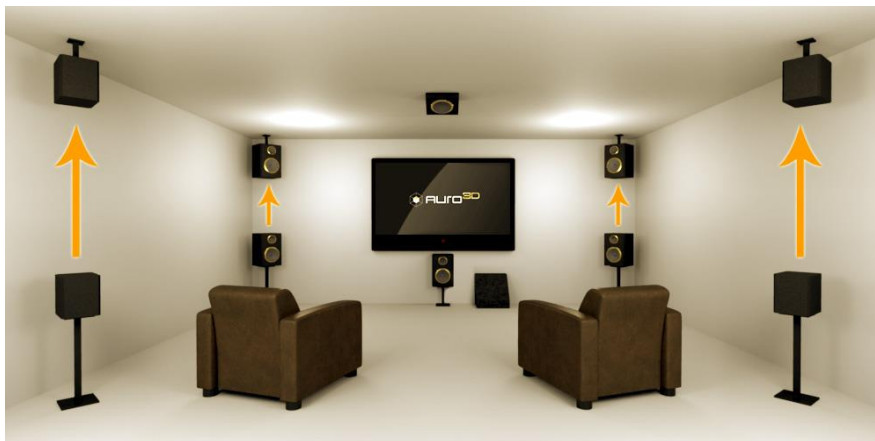
Colocación de los altavoces laterales y superiores en una instalación Dolby Atmos. Fuente: [Dolby](#)

Esta flexibilidad facilitada por los metadatos del sonido basado en objetos, viene acompañada de otras innovaciones, entre las que destacan:

- Altavoces en el eje vertical
- Altavoces con mayor capacidad para reproducir las frecuencias del espectro audible (20 Hz; 20 kHz)
- Presencia de altavoces subwoofers no solo detrás de la pantalla sino también en las zonas traseras y laterales de la sala.
- Gestión específica del rango dinámico para cada entorno de reproducción

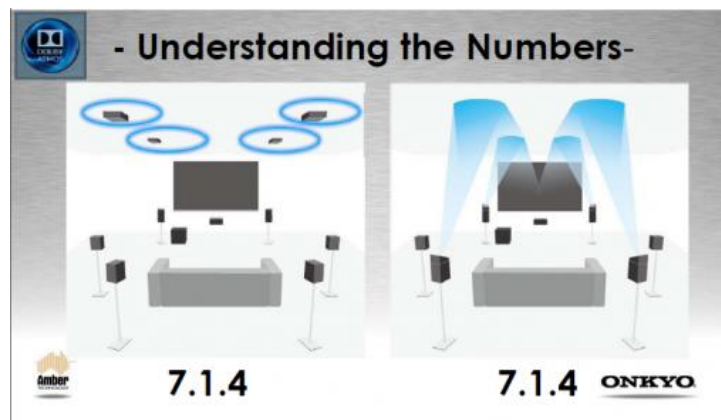
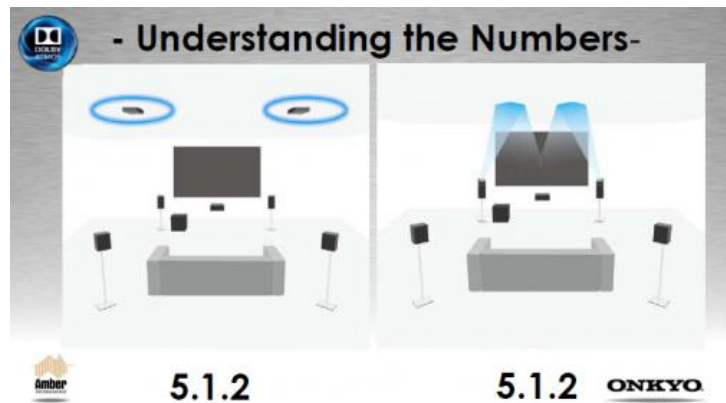
- Mayor precisión de la panoramización de las fuentes de sonido
- Mayor precisión y complejidad de las reverberaciones y delays

El diseño de sonido basado en objetos hace que la experiencia del audio envolvente sea posible con cualquier configuración de altavoces, aunque, como es obvio, a mayor número de altavoces, mejor será la experiencia inmersiva, ya sea en una sala de exhibición o en el hogar (Home Cinema).



Configuración de los altavoces Home Cinema. Fuente: [Auro 3D](#)

Para designar las distintas configuraciones de altavoces se utilizan un código de tres cifras. Por ejemplo, en 9.2.4, la primera cifra (9) indica el número de altavoces tradicionales, la segunda (2), el número de altavoces de subgraves (subwoofer) y la tercera el número máximo de altavoces de altura o eje vertical.



Entendiendo los números: en la derecha con altavoces directos en el techo y a la izquierda utilizando sonido reflejado. Fuente: Onkyo

En cimenatografía hay ahora tres sistemas de sonido inmersivo compitiendo:

- En 2012, Dolby lanzó 'Dolby Atmos' con la película *Brave* (2012, Pixar, Mark Andrews & Brenda Chapman).
- Dos años más tarde, en 2015, DTS entró en escena presentando su sistema de mezcla de sonido por objetos, 'DTS X', de momento más centrado en la electrónica de consumo, como los Blu-ray 4K.
- El tercer actor es 'Auro 11.1', de la compañía Barco, especializada en proyección digital para cine. La primera película que utilizó Auro es *Red Tails* (2012, Lucasfilm, Anthony Hemingway)



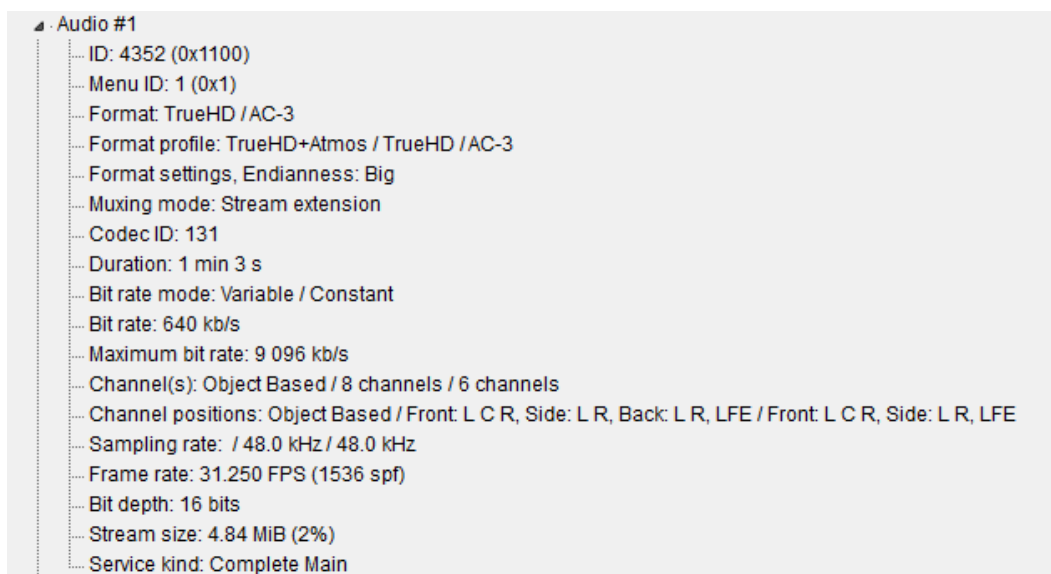
La mayor dificultad para la implantación de estos sistemas inmersivos es la interoperabilidad: cada exhibidor debe apostar y elegir uno para instalarlo en sus salas. Afortunadamente en el ámbito doméstico la mayoría de los componentes (TV UHD, receptores AV, reproductores multimedia...) cuentan con los códecs de varios fabricantes a la vez.

Debido a la diversidad de instalaciones de sonido en las salas de exhibición y los hogares, una misma película se tiene que poder escuchar en estéreo, en 5.1, en 7.1 y en sonido inmersivo 3D. Estas conversiones se basan en el concepto de downmix.

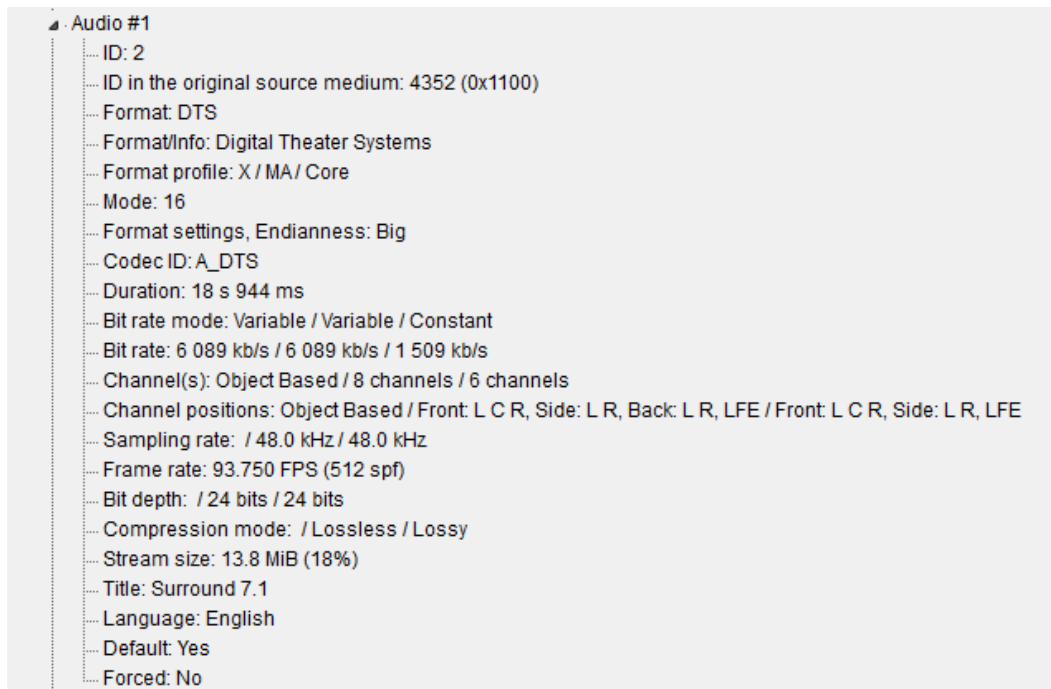
El downmix, es el procedimiento por el cual una mezcla basada en una configuración de más canales/altavoces se reduce a una configuración menor, por ejemplo, pasar de una mezcla 5.1 a un estéreo. No se trata solo de una cuestión de recolocación de las fuentes de sonido, si no también, de delays, coeficientes reductores de volumen, dinámica y ecualización. El proceso de downmix tiene que garantizar que una película se escuchará fielmente en cualquier instalación en un hogar.

Para la adaptación del sonido basado en objetos a los distintos escenarios de recepción domésticos, se utilizan los metadatos que vienen asociados al archivo de sonido MXF.

“El equipo de reproducción, cuando lee el MXF con los metadatos, tiene que detectar el procesador y el modo de distribución, y entonces, hace el ajuste de la espacialidad y la dinámica. Por ejemplo, en la reducción de 64 canales de sonido a 16, hay un conflicto brutal de dinámica. Si no se hace bien, ciertos altavoces pueden acabar saturados. Ese ajuste de espacialidad y ese segundo ajuste de dinámica, es de lo más interesante que traen ‘Dolby Atmos’, o ‘DTS-X’ o ‘Auro11.1’.” (Sergio Márquez)



Metadatos de un archivo MXF preparado para Dolby



Metadatos de un archivo MXF preparado para DTS

Los estudios de sonido certificados por Dolby como 'Dolby Atmos Premier' pueden llegar a tener hasta un máximo de 128 canales mapeados para disponer de 64 outputs simultáneos. Y gracia al hardware con el que están dotados pueden comprobar en tiempo real como queda el resultado en otras configuraciones más reducidas como por ejemplo, 16 altavoces, un entorno 5.1 o incluso un estéreo.

En España todavía hay pocas salas que hayan apostado por el sonido inmersivo, pero afortunadamente, el número va creciendo. Las grandes cadenas, como Kinopolis, Odeon o Warner, han optado por la propuesta de Dolby. 'Auro 11.1' tiene niveles muy bajos de implantación por dos motivos: el primero de ellos es que resulta el menos inmersivo de los formatos contendientes; y el segundo y principal, es que prácticamente solo se instala cuando se ha comprado previamente un proyector DCI de la misma marca, Barco.

Sonido para salas de exhibición de cine

La norma DCI para los servidores DCP es la referencia técnica que se utiliza para la exhibición de cine digital en salas.

Estas son las recomendaciones básicas que hace [Fernando Alfonsin](#) en relación al audio para la creación de un DCP:

- Un DCP puede incluir hasta 32 pistas de audio lineal PCM.
- Las configuraciones más comunes son 5.1, 7.1 y estéreo 2.0.
- Para crear un DCP necesitamos que nos entreguen cada canal por separado en un fichero .WAV (mono) con una resolución de 24 bits y una frecuencia de muestro de 48 kHz o 96 kHz.
- Una vez procesados todos los canales se crea un fichero MXF con ellos.
- El sonido multicanal, al ser PCM, no necesita codificación Dolby y por tanto se ahorra en el coste de esta licencia

Cuando se masteriza una película para DCP se pueden dar las siguientes configuraciones de audio multicanal:

Configuración Canales y formato		
Estéreo	2 PCM	left, right
LCR	3 PCM	left, center, right
5.1	6 PCM	front L., C., front R., surround L, surround R. + Subwoofer
7.1	8 PCM	se añaden los canales traseros: rear L. y rear R.
Immersivo	MXF	sonido por objetos, sonido 3D

El estéreo no se usa en cinematografía. En la configuración mínima se incluye un tercer canal central: LCR (*left, center, right*). Solo se encuentran en estéreo algunas piezas publicitarias o adaptaciones de otros medios que se exhiben también en las salas de exhibición (desgraciadamente sin haber adaptado su mezcla).

El canal central ya se usaba hace décadas, así que en los DCP para exhibición en salas es muy recomendable mezclar incluyendo este tercer canal. Incluso para cortometrajes con presupuestos muy ajustados, la recomendación es por lo menos hacer un LCR.

En largometrajes de ficción o documentales, lo habitual es la configuración 5.1. En España no se está haciendo casi nada en 7.1. Casi todo es 5.1.

El sonido inmersivo / por objetos / 3D, todavía no está muy extendido. Dolby Atmos se está incorporando recientemente, aunque no muchas producciones hacen todavía este tipo de mezcla.



Sala para mezclas de sonido Dolby Atmos Best Digital: fuente Best Digital

Códecs Sonido para televisión en UHD y Blu-ray

Radiodifusión:

En postproducción de sonido para televisión lo ideal es hacer al menos dos mezclas distintas a la original de las salas de exhibición: un 5.1 nuevo y un estéreo, ajustando dinámicas, ecualizaciones etc.

En radiodifusión no es posible emitir a la vez la versión original y la doblada en 5.1 y estéreo. Ocuparía demasiado ancho de banda. Lo habitual es emitir con Dolby Digital y utilizar el sistema automático de downmix para sacar la versión estéreo en el receptor doméstico. Desgraciadamente se produce una pérdida de fidelidad con la mezcla original puesto que el resultado es dependiente de la configuración de receptor que descodifica en el hogar. La mezcla estéreo producida en postproducción siempre estará mejor que el estéreo que se consigue a partir de un downmix de Dolby Digital 5.1.

Para la televisión y la difusión digital de audio multicanal, Dolby creó ‘Dolby Digital Plus’, un códec con más capacidad que el tradicional ‘Dolby Digital (AC-3)’.

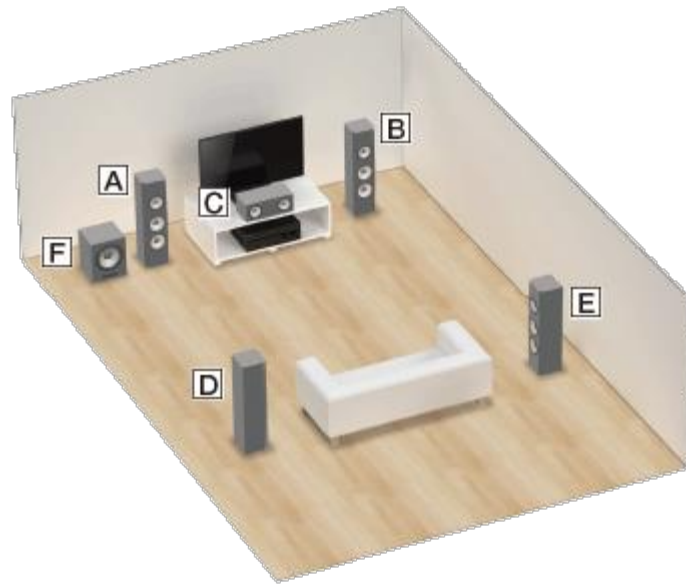


La arquitectura del sistema funciona con un core ‘Dolby Digital’ y unas extensiones (substream) que le van añadiendo más capacidades. ‘Dolby Digital Plus’ puede llegar hasta 6 Mb/s y la extensión ‘Atmos’ ya permite la recepción en hogar del sonido inmersivo.

El core ‘Dolby Digital’ siempre hace falta, porque las extensiones ‘Plus’ y ‘Atmos’ no son reproducibles de forma aislada. Siempre necesitan el paquete anterior, por eso se llaman extensiones.

El ‘Dolby Atmos’ para broadcast, es una versión reducida, pero mantiene el espíritu de la mezcla Atmos que se hizo para los cines. Y es una mejora enorme con respecto al paquete anterior, ‘Digital Plus’ a secas, y del core ‘Dolby Digital’.

Para radiodifusión no hay una solución de DTS, ni siquiera aparece como recomendación en las versiones actuales de la norma BT.2020 como ‘Dolby Digital Plus’.



Configuración de altavoces para home cinema 5.1 en el hogar: A) Altavoz delantero (izquierda) B) Altavoz delantero (derecha) C) Altavoz central D) Altavoz envolvente (izquierda) E) Altavoz envolvente (derecha) F) Altavoz potenciador de graves. Fuente: [Sony](#)

Blu-ray:

Para los Blu-ray se sigue el mismo esquema que para la radiodifusión, pero los códecs son distintos porque no se dan las mismas restricciones en relación con el ancho de banda de la radiodifusión. En los discos Blu-ray la dificultad está en que normalmente se incorporan varios idiomas para facilitar la comercialización internacional y no hay espacio de almacenamiento suficiente en el disco para poner sonido envolvente en todas las versiones.

Por lo tanto, para Blu-ray nos encontramos con códecs distintos que los que se usan en televisión o para internet. Para el sonido multicanal se pueden dar tres casos:

- 1) Sonido 5.1 PCM. Sonido sin compresión, sin patente.
- 2) 'DTS-HD Master Audio'. Sonido sin compresión; hasta 24.5 Mbits/s de flujo de transferencia.
- 3) 'Dolby True HD'. Hasta 18 Mbits/s de flujo de transferencia.



Los primeros Blu-ray venían con sonido 5.1 PCM sin comprimir que no está sujeto a ninguna patente. Pero con este modelo no caben en un disco de 50 GB muchas versiones de idiomas, ni todos los equipos son compatibles con él (aunque parezca paradójico).

Poco a poco el sector se ha ido pasando a 'DTS HD Master Audio', que organiza los ficheros para que ocupen menos espacio, pero sin pérdidas de información (actuando de igual manera que un archivo ZIP). 'DTS HD Master Audio' se utiliza como portador de una extensión que puede llevar el sonido inmersivo DTS-X. Del mismo modo, 'Dolby True HD', permite integrar la extensión Atmos.

Con todo esto, lo habitual en los lanzamientos de Blu-ray 4K de películas, es encontrar este escenario: un 7.1 con DTS-X inmersivo en la versión original; un 7.1 en castellano en DTS HD Master audio; y los demás idiomas en 5.1 o estéreo en Dolby Digital.

En esta imagen se puede ver la contraportada de un Blu-ray 4K donde se ven los diferentes códecs de audio que han incorporado en este lanzamiento comercial.



Logotipos de sistemas de audio en la contraportada de un Blu-ray 4K

Códecs de sonido para difusión por internet

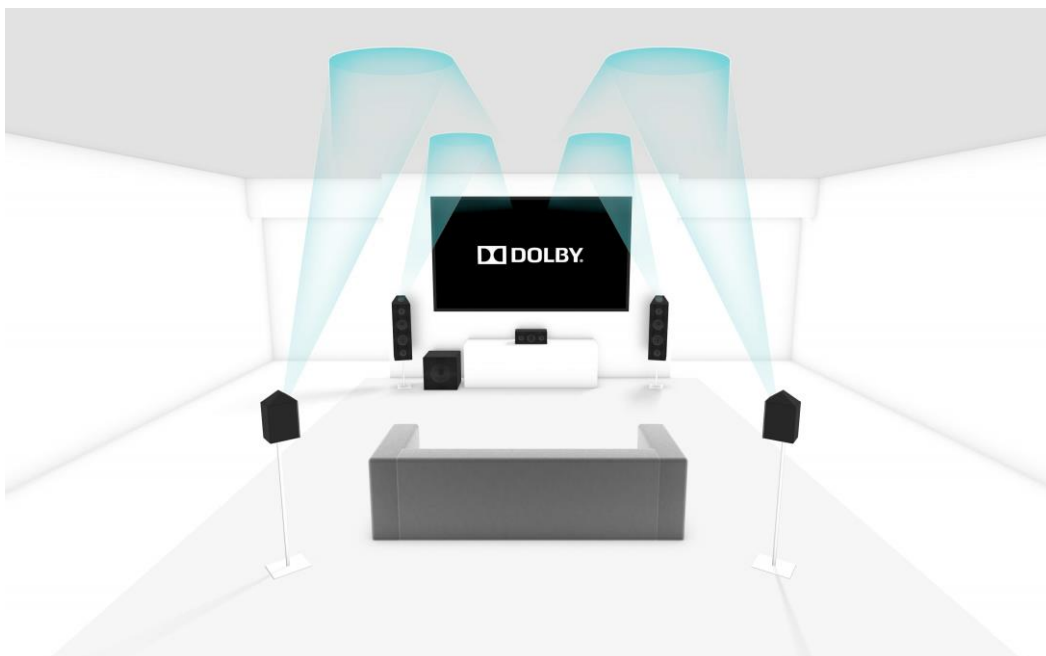
Advanced Audio Coding (AAC) es el formato de sonido que más se está utilizando para difusión de contenidos audiovisuales por internet. Aunque es difícil generalizar, puesto que hay muchos escenarios tecnológicos distintos: V.O.D. través del proveedor de telefonía, televisión por internet OTT/HBBTV, video en páginas web streaming/downloading, etc.

En Internet el factor determinante es la limitación de ancho de banda. En ese sentido es un escenario similar al de la radiodifusión, donde lo que prima son flujos de transferencia muy bajos.

En entornos de vídeo bajo de demanda de pago, el 5.1 está bastante normalizado. Pero se está lanzando el sonido a menos de 320 Kbits/s que es muy poco para que seis canales de sonido suenen fielmente a la mezcla original.

Recientemente con la pujanza de los nuevos actores como Netflix, Amazon Prime, HBO, etc., se están subiendo los flujos de transferencia y utilizando nuevos códecs como 'Dolby Digital Plus' entre 320 y 640 Kbits/s; e incluso ocasionalmente con la extensión Atmos (manteniendo el sonido por objetos, pero en una versión más reducida que la original).

Pese a esto, hay casos sorprendentes, como que Youtube siga sonando sólo en estéreo ¡sin ninguna opción de 5.1!



Configuración de altavoces en el hogar con sonido reflejado en techo. Fuente: Dolby



Altavoz Pioneer S-BS73A para proyectan sonido hacia el techo. Pioneer

Conclusiones

La mejora de la experiencia de usuario de la ultra alta definición tiene que pasar forzosamente por un sonido de mayor calidad. Sin embargo, los desarrollos tecnológicos para mejorar la calidad del audio, que están resueltos desde hace años, no se han implantado de forma generalizada.

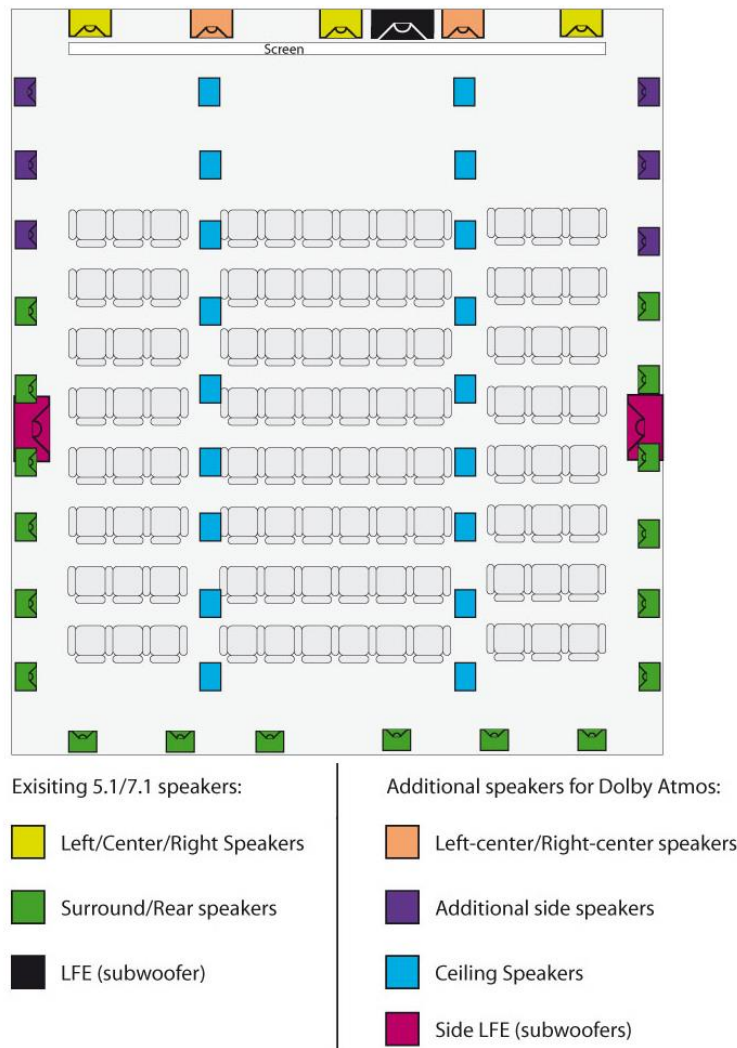
Manuel Sánchez Cid, profesor de la Universidad Rey Juan Carlos (URJC) y experto en sonido envolvente, ha expresado reiteradamente esta idea:

“La ultra alta definición no termina de significarse por incorporar la implementación definitiva de los parámetros de máximo nivel de calidad de sonido logrados hasta la fecha por el desarrollo tecnológico. No obstante, la llegada del Ultra HD posibilita una reapertura hacia a conceptos de inmersión sonora que, sin ser novedosos, se orientan hacia la implementación del segundo plano vertical, así como parece que permite asumir con mayor compromiso una planificación espacial más entroncada con la multiperspectiva y la ruptura del anclaje visual.”

El reto que plantea Sánchez Cid está, por lo tanto, orientado a cuestiones artísticas de realización y a cómo se utilizan los recursos técnicos disponibles. Con los sistemas de sonidos envolvente, son especialmente relevantes la representación del espacio (planos sonoros) y el punto de audición en el que se sitúa al espectador (perspectiva).

Los niveles de calidad que la tecnología de sonido ofrece actualmente son enormes y lo que se está utilizando en la práctica es muy poco en comparación con de lo que se podría hacer.

Sonido en salas de exhibición. En los complejos de exhibición, la tendencia es la actualización o adaptación al sonido inmersivo o 3D en una configuración básica sólo en la sala con mayor capacidad de público. En las salas o cines nuevos, la tendencia es construir aquellas con más capacidad de espectadores directamente en sonido inmersivo en configuración avanzada y el resto o no dotarlas de sonido inmersivo o hacerlo en una configuración pequeña.



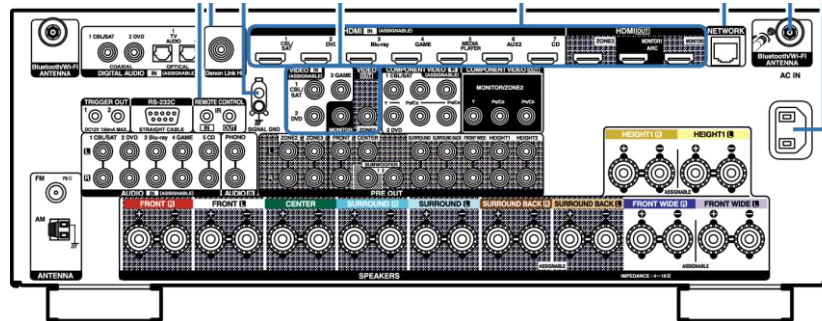
Distribución de altavoces de sonido envolvente para una sala de exhibición de cine.

Radiodifusión. El cuello de botella en cuanto a la calidad de sonido está aquí. Para la captación y la postproducción de sonido hay estándares y herramientas de muy alta capacidad que ven frenada su expansión por la gestión del ancho de banda, el coste económico renovando equipos y la falta de innovación como inversión en imagen corporativa. Las emisiones en mono o estéreo son mayoritarias en la TDT. En la TV por satélite/cable, el Dolby 5.1 convive con producciones estéreo.

Vídeo bajo demanda. Es el actual motor del desarrollo e innovaciones técnicas en la entrega de materiales con mayor calidad de imagen y de sonido, con bit rates más elevados, códecs de última generación, sonido inmersivo etc.

Consumo audiovisual de alta calidad en el hogar o Home Cinema. Los equipos de Home Cinema facilitan las condiciones necesarias para disfrutar el sonido prácticamente con la misma calidad que en el estudio de producción (evidentemente, sin contar el acondicionamiento acústico etc.). Si a estos equipos les llegan señales bajo demanda

online o discos Blu-ray, tenemos al alcance contenidos HDR / 4K / sonido inmersivo/ audio sin pérdidas etc. a todo al máximo nivel, cuando hace unos pocos años esto era impensable en un hogar.



Panel trasero del receptor AV Denon AVR-X4200W para sonido envolvente. Fuente: [Denon](#)